

**Alineación forzada de un corpus de hispanohablantes mexicanos aprendientes de inglés como L2 (COHMAI)**  
*Oral corpus of Spanish speakers learning English as L2 forced alignment*

Esthela Ramírez Hernández  
Universidad Autónoma del Estado de México, México  
lila565.erh@gmail.com

Original recibido: 04/12/22  
Dictamen enviado: 14/06/23  
Aceptado: 28/08/23

**Resumen**

La conformación de corpus orales especializados de aprendientes conlleva una serie de retos metodológicos para su diseño y etiquetación. Investigadores y teóricos sostienen que la segmentación fonética automática es necesaria para la investigación basada en corpus, pues la segmentación manual requiere de mayor tiempo, es costosa e inconsistente. De acuerdo con los resultados, los sistemas de alineación pueden ayudar al estudio automatizado del interlenguaje mediante la generación de corpus alineados. El propósito de este artículo es describir la aplicación de la alineación forzada para la caracterización de la estructura formántica de la vocal epentética. Los datos obtenidos constituyen una primera caracterización acústica de la vocal epentética en aprendientes mexicanos del inglés como L2, y representa un punto de partida de una investigación más amplia en cuanto al número de hablantes, nivel de competencia en la L2 y la variedad de estilos de habla. Los resultados demuestran que vocal epentética es más cerrada que el alófono del español [ɛ].

**Palabras clave:** alineadores, corpus *ad hoc*, inicios consonánticos, interfonología

**Abstract**

*The compilation of L2 learner oral corpora entails a series of methodological challenges regarding both their design and annotation. Researchers argue that automatic phonetic segmentation is a necessary step in their creation, since manual segmentation can be time consuming, expensive, and inconsistent. Furthermore, previous results have shown that forced alignment algorithms can be valuable tools for the corpus-based*

*study of interlanguage. This article describes the use of forced alignment software for the characterization of the formant structure of epenthetic vowels in Mexican learners of English as L2. The collected oral data, and its subsequent analysis, constitutes the first acoustic characterization of epenthetic vowels in the population of interest. The obtained results constitute a starting point for further inquiry considering a higher number of speakers, other L2 proficiency levels, and different speech styles. Results show that epenthetic vowel is closer than the Spanish allophone [ɛ].*

*Keywords: ad-hoc corpus, aligners, initial clusters, interphonology*

## **Introducción**

Los corpus, cualesquiera que sean sus tipos, son esenciales para investigar el uso que se hace de la lengua. El corpus sobre el que se ha trabajado en esta ocasión se caracteriza por ser oral de aprendientes alineado para propósitos específicos. Se diseñó con el objetivo de indagar acerca de la vocal epentética que ocurre como mecanismo de reparación en la producción de inicios consonánticos del tipo sC- (en palabras como *spot*, *stop*, *skill*, *small*, *snap*, *sloop*) en aprendientes del inglés como L2. Se reserva el término “corpus oral” para los corpus orientados a la descripción fonética de la lengua o para el desarrollo de sistemas de reconocimiento o síntesis en el ámbito de las tecnologías del habla (Villayandre Llamazares, 2008, p. 343).

El corpus contiene 4 320 palabras de las cuales 1 080 fueron obtenidas del instrumento de lectura de palabras, 1 080 del instrumento de frases marco, 360 del instrumento de lectura de texto y 1 800 del instrumento de palabras en español. Se emplearon 4 instrumentos de elicitación, los cuales se detallan más adelante en el diseño metodológico. De acuerdo con Hincapié (2018), un corpus de aprendientes se define como:

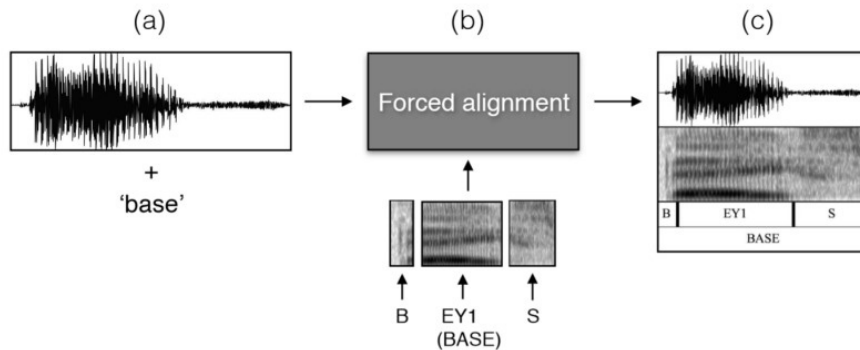
Aquel que reúne muestras de lengua producidas por personas que se encuentran en el proceso de adquisición/aprendizaje de una lengua distinta a la materna. Estas muestras pueden ser de naturaleza oral o escrita y se encuentran sistematizadas, de manera tal que sus metadatos o información adicional permitan la realización de búsquedas y el cruce de información sobre el hablante y la producción lingüística (p. 132).

Investigadores y teóricos especialistas en corpus coinciden en que el tiempo de procesamiento es un gran inconveniente para el etiquetado manual, especialmente cuando se trata de corpus orales (Coto-Solano y Solorzano, 2016; Yuan

y Liberman, 2009; Goldman, 2011). “La ventaja de la alineación forzada es que puede aliviar a un anotador humano de una tarea tediosa” (MacKenzie y Turton, 2020, p. 5). La segmentación automática es el resultado de la alineación forzada, técnica desarrollada para el entrenamiento de sistemas de reconocimiento de voz. Estas consideraciones llevaron a la investigadora a incluir la alineación forzada como parte de la metodología del etiquetado de un corpus oral *ad hoc* de aprendientes, con el objetivo de realizar una caracterización formántica de la vocal epentética en la interlengua español-inglés.

### Sistemas de alineación forzada

La mayoría de los sistemas de alineación forzada se basan en el kit de herramientas de reconocimiento de voz HTK (por sus siglas en inglés *Hidden Markov Model Toolkit*), desarrollado en Cambridge a fines de la década de 1980. Los modelos ocultos de Markov aprenden ventanas que contienen las características auditivas y espectrales de cada fono, por ejemplo, la altura de los formantes y la intensidad de la onda. Después del entrenamiento, estas ventanas son usadas para recorrer la señal de audio y encontrar los límites potenciales entre los fonos (Yuan *et al.*, 2013, p. 1). En la Figura 1 podemos observar la representación esquemática del proceso de la alineación forzada.



**Figura 1.** Representación esquemática del proceso de la alineación forzada en la palabra “base”, que incluye (a) la entrada del audio junto con una transcripción ortográfica, (b) el proceso de alineación forzada, apoyado en un diccionario de pronunciación y en modelos acústicos apropiados, y (c) la salida de un TextGrid alineado en el tiempo con niveles de palabra y de sonido.

Fuente: Bailey (2016, p. 12). Traducción del autor.

Los corpus alineados se utilizan ampliamente en diversas aplicaciones del habla, incluido el reconocimiento automático del habla, la síntesis del habla, así como la investigación prosódica y fonética (Goldman, 2011, p. 1). La alineación forzada se ha empleado en la tecnología de reconocimiento del habla desde los años

setenta y en el área de la investigación de fonética a partir de los años noventa. Las herramientas de alineación forzada que se emplearon en el análisis del presente corpus fueron dos. Para los instrumentos de palabras en inglés, frases marco y lectura de texto, se empleó el alineador forzado del laboratorio de fonética de Penn *P2fa* (Yuan y Liberman, 2009). Se trata de un sistema de alineación entrenado en la Universidad de Pennsylvania que cuenta con modelos acústicos del inglés americano de grabaciones de la Suprema Corte de Justicia. Este alineador se instala como un complemento de Praat y eso hace posible segmentar automáticamente grabaciones, siempre y cuando se provea de un archivo de audio y una transcripción ortográfica. Por otro lado, para analizar el instrumento de palabras del español, se empleó el alineador forzado *faseAlign* (por sus siglas en inglés *Forced Aligned System for Español*) diseñado por Wilbanks (2022). Para realizar la alineación forzada automática de texto a audio, utiliza también el kit de herramientas de reconocimiento de voz HTK. Es importante mencionar que los modelos acústicos del alineador son apropiados para muchas variedades de español latinoamericano, pues, a partir del año 2008, está entrenado con 240 entrevistas y se contemplan el español mexicano, el colombiano y el puertorriqueño.

### Referentes teóricos

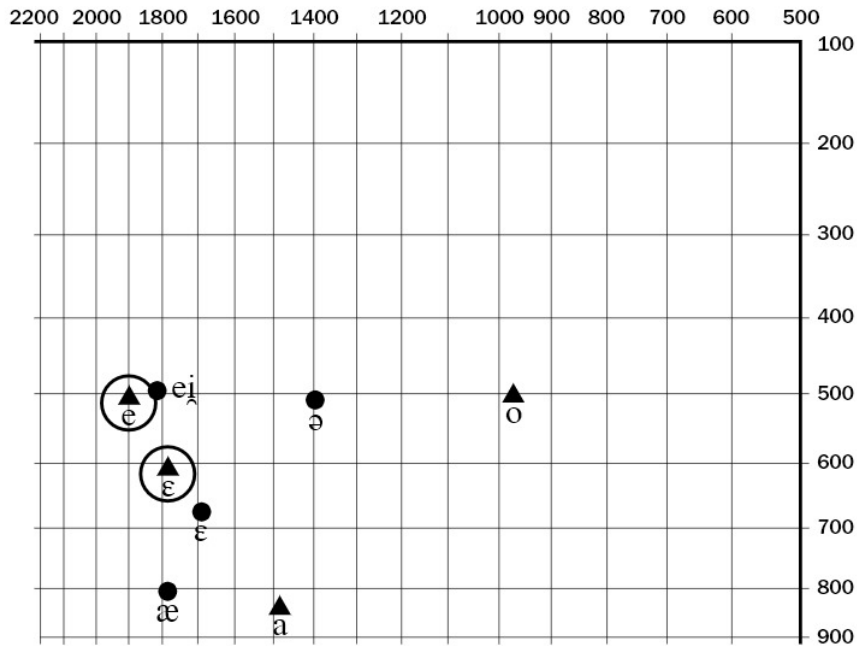
En el siguiente apartado se ofrecen detalles acerca de la vocal media anterior del español /e/, la epéntesis en español y en la interlengua, así como de los estudios previos que han caracterizado acústicamente a la vocal epentética en las interlenguas.

#### *La vocal media anterior del español /e/*

El fonema /e/ es el más frecuente del español con un porcentaje de frecuencia de 14,8 % del total de fonemas. El alófono [e] es el principal del fonema /e/ mientras que [ɛ] es el segundo alófono del fonema. Se trata del segundo porque es el alófono que ocurre en sílabas cerradas, las cuales son menos frecuentes que las sílabas abiertas en español. “El timbre consiste en un primer formante alrededor de 762 Hz. y un segundo formante alrededor de 1300 Hz”. (Clegg y Fails, 2017, pp. 197-199). En la Figura 2 podemos observar contrastes importantes entre los valores formánticos de los dos alófonos de la vocal media del español /e/, a decir [e] y [ɛ], así como de las vocales [a] y [o] y de las vocales del inglés [æ], [ɛ] y [o].

Por otro lado, en el estudio de Marín y Madrid (2001), se obtuvieron los valores formánticos de las cinco vocales del español elicitadas a través frases marco en hablantes femeninos y masculinos de la ciudad de México. En el estudio

mencionado se reportó para vocal /e/ en sílabas cerradas un F1 de 540 Hz. y un F2 de 2 296 Hz. Estos valores nos sirvieron como punto de referencia al momento de comparar los valores formánticos de la vocal epentética de la interlengua español-inglés.



**Figura 2.** “Posición de las vocales [e, ɛ, a, y o] del español (indicadas con un triángulo) y las vocales [æ, ɛ y o] del inglés, así como el punto de partida del diptongo [eɪ] del inglés (indicadas por un punto), según sus formantes. Se indican las vocales objetivo [e] y [ɛ] del español con un círculo”. Fuente: Tomada de Clegg y Fails (2017, p. 200)

### *La epéntesis*

Dentro del marco de teoría de optimidad se explicitan dos tipos: la epéntesis coloreada por el contexto y la epéntesis mínimamente marcada. En el primer tipo, el segmento epentético toma propiedades del segmento siguiente. Por otro lado, la epéntesis mínimamente marcada es aquella en la que se escoge un segmento vocálico no marcado por su punto de articulación (Kager, 2004, p. 124). En los segmentos epentéticos mínimamente marcados, la cualidad está determinada por lo que es más común entre las lenguas, por lo que para epéntesis vocálica las lenguas prefieren las vocales [i], [ə], mientras que para epéntesis consonántica optan por la glotal [ʔ] o la alveolar [t], con el lugar de articulación menos marcado (coronal) y en ocasiones la semivocal [j] (Kager, 2004, p. 125).

Ahora bien, a continuación, analizaremos cómo ocurre la epéntesis vocálica en español con el objetivo de contrastar este mecanismo de reparación con los mecanismos que ocurren en la interlengua, específicamente en la producción de inicios consonánticos del inglés. Al respecto surgen preguntas sobre qué es lo que se prefiere universalmente en la reparación de estos inicios. ¿Se prefiere la epéntesis o la elisión? Paradis y Le Charité (1997) proponen el principio de preservación, cuya premisa sostiene que se debe conservar el material de la estructura profunda a menos que resulte demasiado costoso de mantener para el sistema lingüístico. Generalmente, las lenguas se decantan por la epéntesis como mecanismo de reparación. La elisión también ocurre, pero este mecanismo se presenta mayoritariamente en las lenguas criollas. Otras lenguas como las polinesias combinan el mecanismo de la elisión con el de la epéntesis (Kang, 2008, p. 13).

### *La epéntesis vocálica en el español*

El fenómeno de la epéntesis vocálica del español es recurrente y la vocal que elige el sistema es la /e/ por resultar menos marcada. En español, la epéntesis del plural (farol-*e*-s) y del diminutivo (pan-*e*-cito) no se originan en esencia para reparar estructuras silábicas anómalas, sino para crear una estructura de superficie menos marcada bajo las condiciones de la morfología (Colina, 2003). Una diferencia entre las epéntesis en posición de inicio y las epéntesis en posición final, remarcada por Colina (2009), es que mientras las epéntesis iniciales son sistemáticas e inevitables (aplican a nuevos préstamos), las finales no son productivas. Pero ¿cómo afecta este fenómeno a la adquisición de L2? A continuación, describiremos esta problemática.

### *La epéntesis vocálica en la interlengua*

La epéntesis vocálica es el mecanismo de reparación al que recurren los aprendientes del inglés como L2. Es debido a que subsana la estructura silábica que se altera por influencia de la L2. Tarone *et al.* (1983) caracterizan la epéntesis como una de las estrategias más frecuentes en los sistemas de interfonología. Un fenómeno parecido se da en la adaptación de préstamos que se inician con el inicio consonántico /st/. Por ejemplo, la palabra *stress* [st.ɪ.ɛs] del inglés se emplea en español con una *e* inicial insertada, *estrés* [es.tɾes]. A pesar de que los fenómenos de la interlengua y el préstamo de palabras no son lo mismo, están relacionados debido a que ambos son resultado del contacto entre lenguas. De hecho, muchos préstamos surgen de las formas de la interlengua. Finalmente, se trata de sistemas que intentan simplificar las estructuras consonánticas complejas.

Así, existen pocos estudios que han intentado caracterizar el timbre de las vocales epentéticas en las interlenguas. Por ejemplo, Sandes y Llisterri (2008) realizaron un estudio en el que analizan las vocales epentéticas de la interlengua portugués-español. Los investigadores eligieron a cuatro estudiantes femeninas, hablantes de la variedad del portugués del centro de São Paulo, con un nivel intermedio de conocimiento del español. El corpus analizado se compone de 32 palabras en las que aparecen grupos consonánticos heterosilábicos formados por dos oclusivas orales (/pt/, /bb/, /bt/, /bd/, /db/, /dk/, /kt/), por una oclusiva oral y una fricativa (/bs/, /ds/) y por una oclusiva seguida de nasal (/pn/, /bn/, /tm/, /tn/, /dm/, /gn/). Se seleccionaron 32 palabras que se incluyeron en frases marco del tipo “Dijo ‘aptitud’ y se marchó”, a las que se añadieron 28 frases distractoras. Cada estudiante leyó dos veces la totalidad del corpus. Los resultados del estudio reportan una media de los formantes de las vocales epentéticas analizadas de 367 Hz. para el F1 y de 2 194 Hz. para el F2. Los investigadores concluyen que la vocal epentética tiende a ser más abierta y posterior que [i] y más cerrada que [e], a la vez que presenta una tendencia a la centralización.

## **Diseño metodológico**

### *Selección de la muestra*

Una característica sustancial de los corpus es que estos muestren representatividad de la variedad de lengua que se está investigando. “Al tratarse de una lengua que no es la lengua materna, la variedad de lengua incluye los distintos niveles de dominio que implica su adquisición.” (Mateos, 2012, p. 186). Teniendo en cuenta lo anterior, se determinó que el Marco Común Europeo de Referencia (MCER) podría otorgarnos una distribución por niveles aceptada en la academia, por lo que se buscaron aprendientes de inglés como L2 de los distintos niveles contenidos en el MCER, desde los básicos (A1 y A2) hasta los intermedios (B1 y B2). El total de la muestra quedó conformado por 20 informantes mujeres. El hecho de que sólo participaran alumnos de nivel básico e intermedio se debió a que eran pocos los alumnos que tenían nivel avanzado con las características específicas que buscaba la investigadora; para la selección de la muestra se buscó que las participantes seleccionadas fuesen hablantes nativas de español, nacidas en el Estado de México, y que hubiesen estudiado inglés como L2 en contexto de enseñanza formal. Así, el rango de edad de las informantes oscilaba entre los 17 y 25 años. Todas ellas pertenecen a una clase social media. Ninguna contaba con vínculos étnicos con la L2, ni habían estado en países anglófonos por más de tres meses. Tampoco habían tomado cursos extracurriculares en la lengua meta y nunca habían viajado

al extranjero, por lo que su conocimiento del inglés provenía esencialmente de la instrucción formal pública. Finalmente, la razón por la cual se conformó la muestra exclusivamente de mujeres en esta investigación es el hecho de que dentro de la población estudiantil de la facultad de lenguas predominan estudiantes del sexo femenino.

### Metadatos

Con el objetivo de obtener datos que nos pudiesen arrojar un perfil completo de los informantes, se diseñó y se aplicó un cuestionario sociolingüístico que sirvió para evaluar la presencia de la L2 en el entorno del hablante. A continuación, en el Cuadro 1 se describen las temáticas que se contemplaron dentro de cada factor de análisis del cuestionario sociolingüístico.

CUADRO 1. FACTORES DE ANÁLISIS DEL CUESTIONARIO SOCIOLINGÜÍSTICO.

FACTORES	ASPECTOS GENERALES
Datos de identificación y socioeconómicos	<p>Al tratarse de un estudio de análisis acústico se tuvieron que controlar algunas variables como la clase social.</p> <p>Se decidió trabajar con informantes de clase social media dado que, en un estudio exploratorio de las vocales medias del español en el Valle de Toluca, las producciones de los informantes de esta clase se presentaron más estables y menos sensibles a procesos de debilitamiento (Sánchez, 2014, p. 75).</p> <p>Debido a que el COHMAI (Corpus oral de hispanohablantes mexicanos aprendientes del inglés) incluye un instrumento de elicitación de palabras en español, se contrastaron los valores de los formantes 1 y 2 de la epéntesis vocálica en inicios consonánticos -sC en palabras como “spot” y de las /e/ en posición de inicio en palabras como “España.”</p>
Antecedentes con el aprendizaje del inglés como L2	<p>Para este factor se exploró si la educación básica, media y media superior de las informantes había sido en escuelas privadas o públicas.</p> <p>Se indagó sobre la cantidad de años que llevan estudiando inglés y la edad en la que iniciaron su contacto con este.</p> <p>Se preguntó también por el nivel de inglés con el que ingresaron a la licenciatura y si habían tomado cursos extracurriculares a los tomados en la facultad.</p>
Cantidad de <i>input</i> en horas de inglés: muy alto, alto, medio, bajo y muy bajo.	La cantidad de experiencia con la L2 es una variable que se ha tomado en cuenta por investigadores como Krashen, (1982); Bohn y Flege (1992).
Cantidad de <i>output</i> en horas de inglés: muy alto, alto, medio, bajo y muy bajo.	La cantidad de producción en la L2 es una variable que se ha tomado en cuenta por investigadores como Guion, <i>et al.</i> , (2000); Peñate y Arnaiz, (2004).



La cantidad de horas de *input* y de *output* se conoció a través del cuestionario sociolingüístico mediante la pregunta ¿cuántas horas en una escala que va de cero horas hasta siete horas o más por semana realiza las siguientes actividades? Para la cantidad de *input* se enumeraron cinco actividades principalmente: escuchar a los maestros hablar en inglés, ver videos en *YouTube* sin subtítulos, leer por placer libros, periódicos, escuchar *podcasts*, audios y música, usar aplicaciones o jugar videojuegos en inglés. Por otro lado, respecto a la cantidad de *output* se enumeraron cinco actividades principalmente: redactar proyectos en inglés, dar clases o cantar en inglés, escribir en las redes sociales en inglés, visitar el SAC (*Self Access Center* por sus siglas en inglés) para conversar o escribir en inglés y, finalmente, hablar con amigos y /o familiares en inglés. Así, se diseñó una escala que contiene rangos de puntaje de cero a veinte hasta llegar al puntaje máximo de cien. Además, se establecieron cinco variantes de las variables extralingüísticas: cantidad de *input* y cantidad de *output* (muy baja, baja, media, alta y muy alta). Es relevante mencionar que el porcentaje de *input* fue siempre mayor o igual al porcentaje del *output*. Jamás fue superior, lo cual podría ser indicativo de que las informantes no tienen suficiente interacción con la L2 para poner en práctica lo aprendido en clase.

### *Instrumentos*

Se elicitaron muestras de habla formal por lo que se trabajó con cuatro instrumentos. El objetivo primordial fue obtener habla formal a través de la lectura de palabras, lectura de frases marco, lectura de textos que tuvieran palabras en inglés con inicios consonánticos del tipo: /sp/, /st/, /sk/, /sl/, /sm/, /sn/ y lectura de palabras en español.<sup>1</sup>

Los corpus orales se graban bajo condiciones muy controladas y suelen consistir en segmentos aislados, frases aisladas, textos leídos[...] En general, se diseñan con mucho cuidado para recoger el fenómeno objeto de estudio y tienen un tamaño reducido al no utilizar un número elevado de hablantes [...] (Villayandre Llamazares, 2008, p. 343).

A continuación, en el Cuadro 2, se detalla en qué consistieron cada uno de los instrumentos, así como el tiempo aproximado de duración.

<sup>1</sup> Los estudios de interlengua basados en corpus requieren, por lo tanto, un corpus nativo comparable, esto es, diseñado para permitir comparaciones (Rufat, 2015, p. 197).

CUADRO 2. INSTRUMENTOS DE INVESTIGACIÓN.

INSTRUMENTO	CARACTERÍSTICAS	EJEMPLO	DURACIÓN
Cuestionario sociolingüístico	31 preguntas en <i>Google forms</i>	¿Cuántas horas a la semana escuchas música en inglés?	20 minutos
Lectura de palabras en inglés	18 palabras monosilábicas CCV con inicios consonánticos sC- más 36 palabras distractoras con verbos en pasado regular.	<i>asked</i> <i>complained</i> <i>speech</i> <i>traveled</i> <i>jumped</i>	6 minutos
Lectura de frases marco en inglés	18 frases con palabras metas más 36 frases distractoras con verbos en pasado regular.	<i>Repeat asked naturally</i> <i>Repeat complained naturally</i> <i>Repeat speech naturally</i> <i>Repeat traveled naturally</i> <i>Repeat jumped naturally</i>	10 minutos
Lectura de texto en inglés	6 palabras con inicios consonánticos sC-, después de oclusiva sorda.	(...night <i>small</i> ..)	4 minutos
Lectura de palabras en español	30 palabras en español con VCC más 60 palabras extrañas en español para distraer al informante.	Faroles Contabilidad España Comunes Palatal	15 minutos

### *Grabaciones: aspectos técnicos*

La recolección de muestras se llevó a cabo en un aula digital de la Facultad de Lenguas de la Universidad Autónoma del Estado de México (UAEMex). La grabación se realizó en un espacio aislado. El micrófono se situó a 10 cm del interlocutor. Los archivos de audio de los aprendientes se exportaron y almacenaron en formato WAV. La versión final de los archivos de COHMAI se ofrece en formato WAV. Respecto a las consideraciones técnicas del levantamiento del corpus, las grabaciones se realizaron con el programa *Speech Analyzer* (versión 3.1). Esta herramienta fue desarrollada por SIL (*Summer Institute of Linguistics* por sus siglas en inglés) y sus funciones principales son las de grabar y analizar archivos de sonido. Se eligió por ser un software especializado en la grabación y análisis de habla con el que la investigadora había trabajado en investigaciones previas.

### *Materiales*

Las grabaciones se realizaron con dos laptops, una Dell, en la que se mostraron los instrumentos a los colaboradores y una laptop Hewlett-Packard (HP) en la que se registraron las grabaciones. Asimismo, se empleó un micrófono cardioide *Shure SM58*, también conocido como micrófono unidireccional, pues captura

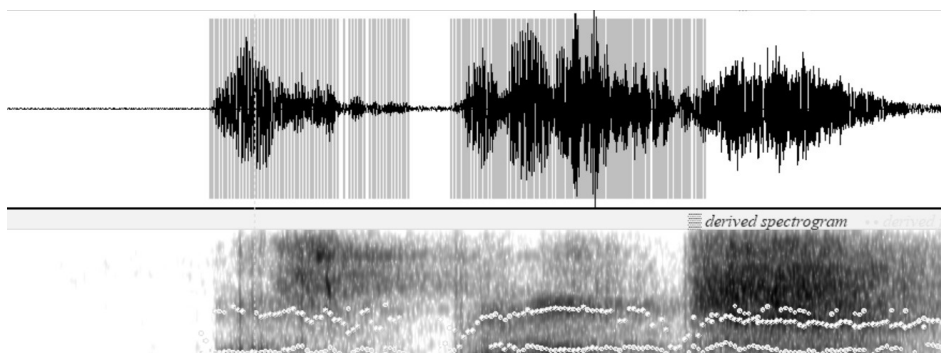
el sonido en una sola dirección, a saber la parte de enfrente de la capsula y sus laterales. Se trabajó con una frecuencia de muestreo de 22.05 KHz.

### *Criterios para el análisis de mecanismos de reparación*

Para determinar la ocurrencia de las vocales epentéticas se consideraron tres criterios acústicos: la intensidad de energía del espectro, la duración del segmento vocálico y la cualidad de la estructura formántica. Se ha procedido a la observación de posibles indicios visuales de la presencia de un sonido vocálico en el oscilograma y en el espectrograma. Así, en la Figura 3 podemos observar la señal acústica de la palabra *speech*. En dicha imagen se destaca la epéntesis vocálica producida por la informante seis.

### *Criterios para la segmentación de las vocales epentéticas*

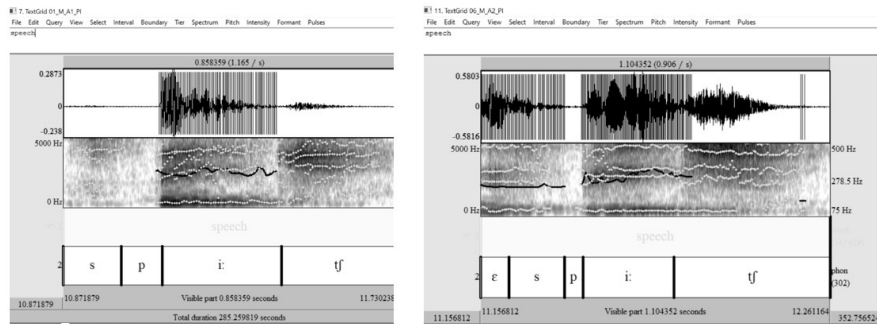
La vocal epentética se trata de un mecanismo de reparación al que recurren los aprendientes del inglés como L2 para subsanar estructuras silábicas marcadas. Para determinar la ocurrencia de las vocales epentéticas se consideraron tres de los criterios acústicos antes mencionados (intensidad de energía del espectro, duración del segmento vocálico y cualidad de la estructura formántica). En la Figura 4, se presenta la comparación del oscilograma y el espectrograma de la pronunciación de la palabra *speech* del instrumento de palabras en inglés, y se observa que la informante dos del COHMAI no recurre a la epéntesis como mecanismo de reparación ante el ataque complejo /s + oclusiva/; mientras que la informante seis al producir la misma palabra sí recurre a la epéntesis en el instrumento de palabras del inglés.



**Figura 3.** Oscilograma y espectrograma de la pronunciación de palabra *speech* ('discurso') con epéntesis vocálica por la informante seis del COHMAI. Elaboración propia.

## Digitalización

En la fase de digitalización se extrajeron las muestras recogidas por el software *Speech Analyzer* para almacenarlas en el ordenador, con el fin de tratarlas y trabajar en ellas con Praat. Este es un *software* libre para el análisis científico del habla usado en lingüística, diseñado por Paul Boersma y David Weenik en la Universidad de Ámsterdam.



**Figura 4.** Comparación de pronunciations de la palabra *speech* ('discurso') sin y con vocal epentética de la informante dos (izquierda) y seis (derecha) del COHMAI.

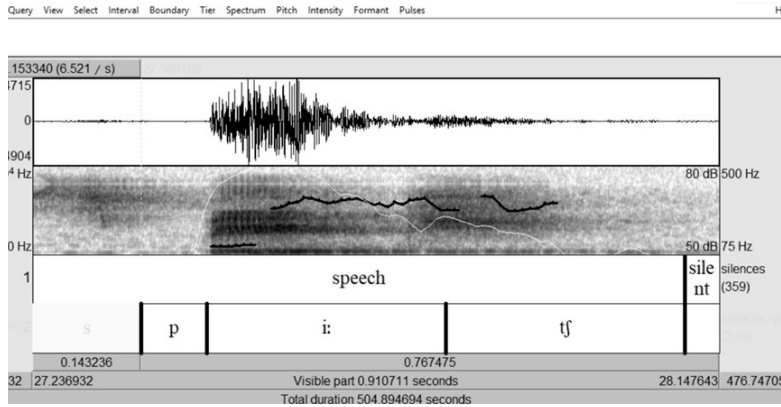
## La alineación forzada con *P2fa*

La alineación para el instrumento de palabras en inglés, frases marco y lectura de texto se realizó con *P2fa* (alineador forzado del laboratorio de fonética de la Universidad de Pennsylvania), por medio del programa Praat. A continuación, se detalla el proceso de preparación de los datos para la alineación forzada. En primer lugar, se eliminaron los ruidos y la información no deseada de las grabaciones. Para etiquetar los silencios y agregar las etiquetas de las palabras en inglés, se emplearon dos *scripts* diseñados por Mietta Lennes de la Universidad de Helsinki: *label\_from\_textfile* y *mark-pauses*.<sup>2</sup> El *script mark-pauses* revisa la frecuencia de la señal de audio y marca los silencios. Todo registro debajo de 59 decibeles es marcado como un silencio. Mientras, el *script label\_from\_textfile* agrega las palabras en inglés al audio. Para alinear las etiquetas con la señal sonora, se seleccionó, en la parte superior de Praat, la pestaña *Interval* y dimos clic en *Alignment Settings* -> *include phonemes only*. En la Figura 5, se observa la producción de la palabra *speech* (dicurso) segmentada a través del alineador *P2fa*.

En la alineación forzada la salida típica de estos sistemas es un *TextGrid*, un archivo de texto (Boersma y Weenink, 2018), donde cada palabra y fono identificado en el audio se empareja con su transcripción. Posteriormente, se ejecutó

<sup>2</sup> <https://lennes.github.io/spect/>

un *script* de análisis de datos. Después de haber corrido el *script* de análisis automático, se obtuvo una base de datos en la cual las filas contienen datos y las columnas variables. Dentro de la base de datos se anotó, en la fila superior, la glosa, la transcripción fonológica y la fonética, y la frecuencia del formante 1 y 2 de la vocal epentética.



**Figura 5.** Alineación forzada de la palabra *speech* ('discurso') de la informante dos.

### *La alineación forzada con fase Align*

A continuación, se detalla el proceso de preparación de los datos en español para la alineación forzada. En primer lugar se eliminaron los ruidos y luego se prescindió de la información no deseada en las grabaciones. Una vez finalizado el proceso de limpieza se procedió a la tarea de la transcripción fonológica a través de la alineación forzada. Para etiquetar los silencios y agregar las etiquetas de las palabras en español se emplearon los mencionados *scripts* de Lennes (2001). Una vez etiquetado el archivo se procedió a alinearlos. Es importante mencionar que *faseAlign* alinea exclusivamente lo que tenga etiquetas, por ello es importante corroborar que se encuentre etiquetado únicamente aquello que se desea alinear. Finalmente, se ejecutó un *script* de análisis de datos. Después de haber corrido el *script* de análisis automático, se obtuvo una base de datos. Dentro de esta, tal y como se había comentado antes, se anotó en la fila superior la glosa, la frecuencia del formante 1 y 2 de la vocal media anterior del español /e/. En la Figura 6, se observa la producción de la palabra “espejo” de la informante seis del COHMAI, segmentada a través del alineador *faseAlign*.

Sin embargo, como lo sugiere Goldman (2011), se recomienda emplear esta metodología con habla no espontánea debido a que la espontánea presenta una considerable cantidad de variaciones, lo cual podría no favorecer el uso de esta herramienta

de alineación forzada. “Incluso con herramientas informáticas y preparación de datos precisos, los sistemas automáticos pueden cometer errores que un ser humano no cometería” (Goldman, 2011. p. 1). Así, MacKenzie y Turton (2020) sugieren incluir en el análisis una etapa de post- procesamiento a través de la confirmación humana, para corregir errores que pueden desplazar tramos completos del discurso y así mejorar el proceso de transcripción y etiquetación.

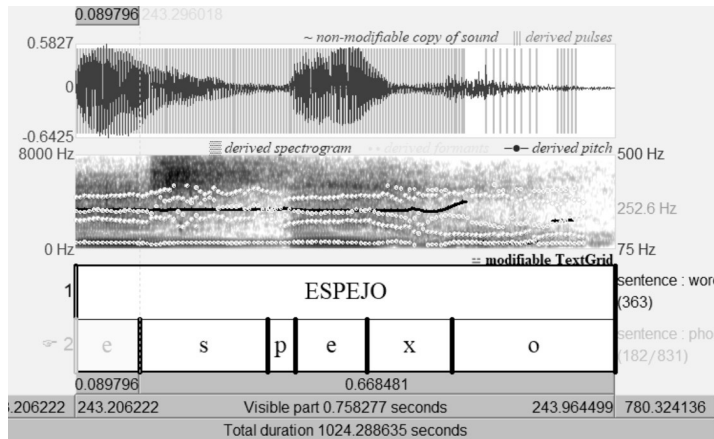


Figura 6. Alineación forzada de la palabra “espejo” [Es'pexo] de la informante seis.

## Análisis y discusión de resultados

El objetivo general de este estudio es la descripción formántica de las vocales epentéticas. A continuación, mostramos resultados de la producción de la vocal epentética de la informante seis del COHMAI. En la Figura 7, se muestra el formante 1 (F1) correlato acústico relacionado con la altura de la lengua dentro de la cavidad oral, y el formante 2 (F2) relacionado con la anterioridad de la lengua. Así, en dicha figura, observamos variación en cuanto a la altura y anterioridad de la lengua, en contraste con los valores formánticos de la vocal /e/ del instrumento de palabras en español. Los resultados reportan que las vocales epentéticas de la informante seis presentan una media del F1 de 402 Hz. y una media en el F2 de 2 087 Hz., valores que se encuentran entre la producción de una /e/ y una /i/ del español; pues en el estudio de Madrid y Marín (2001) se reporta un F1 para /e/ de 584 Hz. y un F2 de 2 296 Hz., mientras que para la /i/ se registra un valor de F1 de 323 Hz. y un F2 de 2 875 Hz.

En la Figura 8, se observan el ploteo de los formantes 1 y 2 de la vocal [ɛ], resultado de la lectura de 30 palabras en español que realizó la informante seis en tres ocasiones. Los resultados indican que la media de las 90 mediciones del F1 es de 511 Hz. y la media del F2 es de 2 381 Hz. Estos resultados son indicadores de

que la vocal [ɛ] en sílaba cerrada tiene una altura y anterioridad muy similar a la reportada en el estudio de Madrid y Marín (2001).

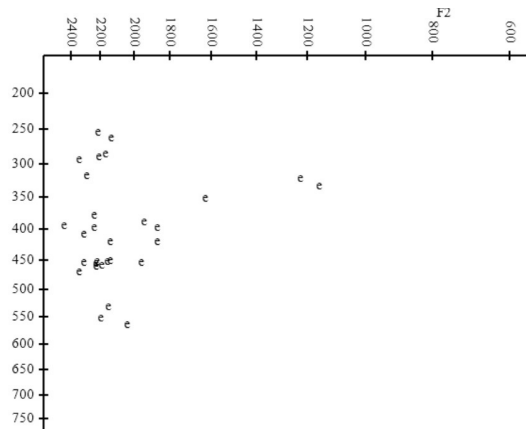


Figura 7. F1 y F2 de la vocal epentética en palabras del inglés de la informante seis del COHMAI.

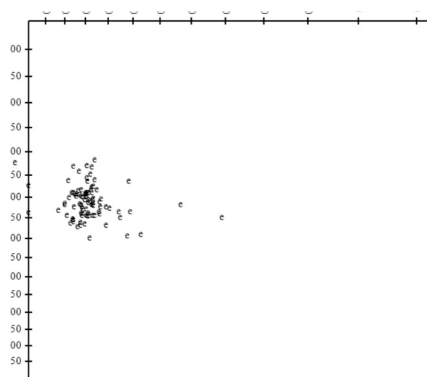
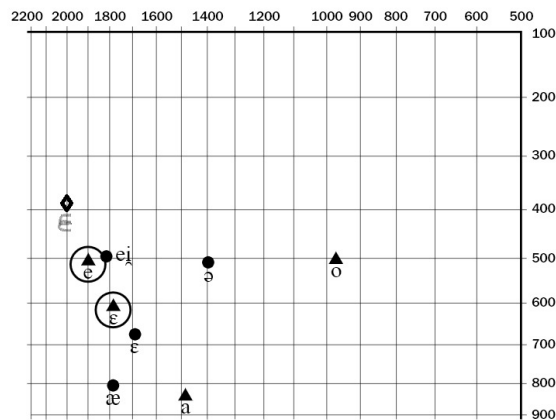


Figura 8. F1 y F2 de la vocal media anterior [ɛ] en posición inicial, en palabras de español con sílaba cerrada de la informante seis.

Después de comparar ambas medias de los formantes, podemos interpretar que hay diferencias entre las medias del F1 y el F2. Lo anterior sugiere que el rasgo de la altura y anterioridad de la lengua se modifica al articular las vocales epentéticas en contraste con las vocales medias anteriores /e/ en posición de inicio y en sílabas cerradas como en la palabra “espejo”. Es decir, la vocal epentética tiende a ser más cerrada en contraste con la /e/ del español. En la Figura 9, ilustramos los valores formánticos de la vocal epentética reportados en el presente estudio, en contraste con valores formánticos de algunas vocales del español y del inglés.

Estos datos arrojan luz al campo de estudio de la fonética acústica en cuanto a la caracterización formántica de las vocales epentéticas de la interlengua español-

inglés pues, de acuerdo con la revisión del estado del arte, existen pocos estudios que analicen específicamente la cualidad espectral de la vocal epentética de las interlenguas. Citamos el estudio de Sandes y Llisterri (2008), quienes caracterizaron acústicamente a la vocal epentética [ɪ] de la interlengua portugués-español. Estos investigadores reportaron que la vocal epentética tiende a ser más abierta y posterior que [i] y más cerrada que [e], por lo que presenta tendencia a la centralización, lo cual es coincidente con los hallazgos del presente estudio.



**Figura 9.** La posición de las vocales [e, ɛ, a y o] del español (indicadas con un triángulo) y las vocales [æ, ɛ y o] del inglés, así como el punto de partida del diptongo [eɪ] del inglés (indicadas por un punto), según sus formantes. Se señalan las vocales objetivo [e] y [ɛ] del español con un círculo y la vocal epentética con un rombo. Fuente: Adaptada de Clegg y Fails (2017, p. 200).

## Conclusiones

Los hallazgos obtenidos constituyen una primera caracterización de la estructura formántica de la vocal epentética. Esto representa un punto de partida de una investigación más amplia en cuanto al número de hablantes, el nivel de competencia en la L2 y la variedad de estilos de habla. Así pues, se sugiere para estudios futuros integrar análisis de la duración de la vocal epentética, dado que se trata de un correlato acústico importante para la caracterización de las vocales epentéticas en las interlenguas.

Por otro lado, es importante señalar que el diseño y elaboración de un corpus oral conlleva una serie de retos metodológicos para el investigador. Por ende, se sugiere el uso de la alineación forzada como herramienta para la realización de estudios en las interlenguas, con lo cual se podría ahorrar tiempo en el proceso de etiquetación.

Finalmente, he de mencionar que este trabajo es meramente exploratorio, de ahí que se sugiera que en investigaciones futuras se contrasten la efectividad de



otros alineadores como *Easy Align* o *Fave Align*. Por otro lado, son necesarios más estudios de corpus orales especializados que empleen la alineación forzada para corroborar estas aportaciones respecto del entrenamiento de modelos acústicos en las interlenguas.

### Agradecimientos

Quisiera agradecer la invaluable ayuda de Max Carey Anderson y de Arturo Curiel Díaz para la aplicación de los alineadores forzados *P2fa* y *faseAlign*, así como el diseño de *scripts* para el análisis de datos. También deseo agradecer cada una de las observaciones y sugerencias puntuales de los tres dictaminadores anónimos de este artículo, quienes invirtieron tiempo en la lectura y el análisis minucioso del texto. Gracias a su exhaustiva revisión pude resolver fallos que la primera versión de este trabajo contenía. Cualquier inconsistencia o error de la presente investigación debe atribuirse exclusivamente al autor.

### Referencias

- Bailey, G. (2016). "Automatic detection of sociolinguistic variation using forced alignment". *University of Pennsylvania Working Papers in Linguistics: Selected Papers from New Ways of Analyzing Variation* (NWAYV 44), pp. 11-20. [https://eprints.whiterose.ac.uk/139456/1/Automatic\\_Detection\\_of\\_Variation.pdf](https://eprints.whiterose.ac.uk/139456/1/Automatic_Detection_of_Variation.pdf)
- Bohn, O. S., & Flege, J. E. (1992). "The production of new and similar vowels by adult German learners of English". *Studies in Second Language Acquisition*, 14, pp. 131-158. <https://doi.org/10.1017/S0272263100010792>
- Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Versión 6.0.40. Consultado el 3 de mayo del 2023 de <http://www.praat.org>
- Clegg, J., y Fails, W. (2017). *Manual de fonética y fonología españolas*. Routledge: Nueva York.
- Coto-Solano, R., y Solórzano, S. F. (2016). "Alineación forzada sin entrenamiento para la anotación automática de corpus orales de las lenguas indígenas de Costa Rica". *Kánina*, 40(4), pp. 175-199. <https://doi.org/10.15517/rk.v40i4.30234>
- Colina, S. (2003). "The status of word-final [e] in Spanish". *Southwest Journal of Linguistics*, 22(1), pp. 87-108.
- Colina, S. (2009). *Spanish phonology: A syllabic perspective*. Georgetown University Press: Washington.
- Goldman, J. P. (2011). "EasyAlign: an automatic phonetic alignment tool under Praat". *Interspeech'11, 12th Annual Conference of the International Speech Com-*

- munication Association*. [https://www.iscaspeech.org/archive\\_v0/archive\\_papers/interspeech\\_2011/i11\\_3233.pdf](https://www.iscaspeech.org/archive_v0/archive_papers/interspeech_2011/i11_3233.pdf)
- Guion, S. G., Flege, J. E., Liu, S. H., y Yeni-Komshian, G. H. (2000). "Age of learning effects on the duration of sentence produced in a second language". *Applied Psycholinguistics*, 21(2), pp. 205-228. [http://jimfege.com/fles/Guion\\_Flege\\_age\\_efects\\_AP\\_2000.pdf](http://jimfege.com/fles/Guion_Flege_age_efects_AP_2000.pdf)
- Hincapié, D. (2018). "Corpus de Aprendientes de Español como Lengua Extranjera y Segunda Lengua (caele/2): el componente escrito". *Forma y Función*, 31(2), pp. 129-143. <https://revistas.unal.edu.co/index.php/formayfuncion/article/view/74659/67643>
- Kager, R. (2004). *Optimality Theory*. Cambridge: Cambridge University Press.
- Kang, Y. (2011). "Loanword phonology". *The Blackwell companion to phonology*, IV, pp. 1-25. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781444335262.wbctp0095>
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press Inc.
- Lennes, M. (2001). SpeCT-The Speech Corpus Toolkit for Praat (previously "Mietta's Praat scripts").
- Llisterri, J. (1991). *Introducción a la fonética: el método experimental*. Barcelona: Anthropos.
- Madrid, E. y Marín, M. (2001) "Estructura formántica de las vocales del español de la ciudad de México". En E. Herrera (Ed.). *Temas de fonética instrumental* (p). México, D. F.: El Colegio de México.
- MacKenzie, L., & Turton, D. (2020). "Assessing the accuracy of existing forced alignment software on varieties of British English". *Linguistics Vanguard*, 6(1), pp. 1-14. <https://doi.org/10.1515/lingvan-2018-0061>
- Mateos, A. V. (2012). *Análisis de errores de aprendientes de francés lengua extranjera (FLE) basado en corpus orales* (Tesis doctoral). Universidad Autónoma de Madrid, Madrid. <https://dialnet.unirioja.es/servlet/dctes?codigo=36228>
- Paradis, C. & D. La Charité. (1997). "Preservation and minimality in loanword adaptation". *Journal of Linguistics*, 33, pp. 379-430. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=481575e149b01fd24fc60bfe5440b7a7a10dfa84>
- Peñate, M. y P. Arnaiz. (2004). "El papel de la producción oral (output) en el proceso de aprendizaje de una lengua extranjera (LE)". *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*, (1), pp. 37-59. <https://dialnet.unirioja.es/servlet/articulo?codigo=1129989>

- Rufat, A. S. (2015). “Análisis contrastivo de interlengua y corpus de aprendientes: precisiones metodológicas”. *Pragmalingüística*, (23), pp. 191-210. <https://core.ac.uk/download/pdf/230887407.pdf>
- Sánchez, J. A. (2014). “Análisis exploratorio de las vocales medias en el español del Valle de Toluca”. *Verbum et lingua: Didáctica, lengua y cultura*, (4), pp. 69-79. <http://verbumetlingua.cucsh.udg.mx/index.php/VerLin/article/view/37>
- Sandes, E. I., y Llisterri, J. (2008). “Estudio acústico de las vocales epentéticas en la interlengua de estudiantes brasileños de E/LE”. *V Congresso Brasileiro de Hispanistas/I Congresso Internacional da Associação Brasileira de Hispanistas*, pp. 2521-2529. [https://www.joaquimllisterri.cat/publicacions/Sandes\\_Llisterri\\_08\\_Vocales\\_Epenteticas\\_ELE.pdf](https://www.joaquimllisterri.cat/publicacions/Sandes_Llisterri_08_Vocales_Epenteticas_ELE.pdf)
- Tarone, E., & Cohen, A. D., Guy, D. (1983). “A closer look at some interlanguage terminology: a framework for communication strategies”. In C. Færch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 4-14). New York: Longman. Recuperado de <https://eric.ed.gov/?id=ED125313>
- Villayandre Llamazares, M. (2008). “Lingüística con corpus” (I). *Estudios Humanísticos. Filología*, (30), pp. 329-349. <https://dialnet.unirioja.es/servlet/articulo?codigo=3332675>
- Wilbanks, E. (2022). faseAlign (Version 1.1.14) [Computer software]. <https://github.com/EricWilbanks/faseAlign>
- Yuan, J. y Liberman M. (2009). “Investigating /l/ variation in English through forced alignment”. En: *Proceedings of InterSpeech*, pp. 2215-2218. [https://www.isca-speech.org/archive/interspeech\\_2009/yuan09\\_interspeech.html](https://www.isca-speech.org/archive/interspeech_2009/yuan09_interspeech.html)